Electronic Supplementary Material (ESI) for CrystEngComm.

Electronic Supplementary Information: Molecular simulation of oligo-glutamates in a calcium rich aqueous solution: insights into peptide-induced polymorph selection

Jens Kahlen[a], Christine Peter[b], Davide Donadio[acd],

[a] *Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany*

[b] *University of Konstanz, Department of Chemistry, 78457 Konstanz, Germany*

[c] *Donostia International Physics Center, Paseo Manuel Lardizabal 4, 20018 San Sebastian, Spain.*

[d] *Ikerbasque, Basque Foundation for Science, Bilbao, Spain.*

(Dated: May 15, 2015)

A. Molecular dynamics simulations: computational details

We have simulated systems consisting of one peptide (5Glu or 10 Glu) with and without Ca ions (at a ratio of 10 Ca^{2+} ions per peptide molecule of 10, in accordance with the experimental conditions [1]). We used the GROMACS simulation package [2, 3] (version 4.6.3) and the GROMOS 54A8 FF [4] with optimised short-range interaction parameters between Ca^{2+} ions and oxygen atoms of the carboxylate group of the Glu side chains [5]. in combination with the standard GROMOS parameters for calcium [6] and with the SPC/E water model [7]. The 5Glu systems contained 3563 or 3538 SPC/E [7] water molecules ($\sim$108 nm^3 box), the 10Glu systems contained 8203 or 8183 water molecules ($\sim$249 nm^3 box) and were neutralized with an appropriate number of $Cl-1$ ions. All simulations were conducted at 300 K (maintained *via* stochastic velocity rescaling [8] with a coupling time of 0.1 ps). The equations of motion were integrated with the leap-frog algorithm with a time step of 2 fs. The center of mass translation of the simulation box was removed every 200 fs and the neighbour list was updated every 20 fs. The particle mesh Ewald method [9] was used for long-range electrostatic interactions, with a grid spacing of 0.1 nm and an interpolation of 4. Lennard-Jones interactions and real-space Coulomb interactions were truncated at 1.4 nm, no long-range dispersion correction was applied.

B. H-REMD simulations: the biasing potential

The dihedral angle BP-REMD method is a special variant of the H-REMD method, which has been designed specifically for the enhancement of the conformational sampling of peptides and proteins [10, 11]. The conformations of peptide backbones are generally characterised by one pair of dihedral angles φ and ψ per amino acid monomer. In order to enhance the conformational sampling in these systems, a biasing potential is introduced in BP-REMD simulations, which enforces changes in the dihedral angles φ and ψ along the backbone. The level of biasing - and with it the frequency of conformational transitions - is gradually increased along the replicas. This ensured that the system can escape from getting trapped in local energy minima.

To set up the biasing-potential, we determined the potentials of mean force (PMF) of the two backbone dihedral angles φ and ψ. The PMFs of φ and ψ of the mid-monomer of

a pentamer of glutamate were determined with well-tempered metadynamics [12] using the PLUMED2 [13] plug-in for GROMACS. The settings during these simulations were chosen in accordance with those of Barducci *et al.*, who have studied the free energy of rotation around the dihedral angles of an alanine dipeptide: The height of the Gaussians in the beginning was set to 1.2 kJ/mol, the deposition interval was 120 fs, a bias factor of 5 was used and the width of the Gaussian hills was set to 0.1745 rad corresponding to 10 °. For each of the two dihedral angles, the peptide solvated in 3563 water molecules was simulated in a dodecahedron box in the NVT ensemble for 4 ns. The PMFs were determined in the absence of any calcium ions.

For use with the GROMACS package [2, 3], the biasing-potentials were fitted to the following functional form:

$$V(\varphi) \;=\; \sum_{i=1}^{6} k_i \cdot [1 + \cos(n_i \varphi - \delta_i)] \tag{1}$$

Here, the multiplicities n_i are set to $n_1 = 1$, $n_2 = 2$, etc. The same functional form is used for the dihedral angle ψ. The results of this optimisation procedure is given in table I, the fitted potentials are shown in figure 1 together with the original PMFs.

TABLE I: Parameter values of the curves that has been fitted to the PMFs of the dihedral angles φ and ψ. The functional form of the fitted curves (eq. 1) corresponds to form of the peptide backbone dihedral potentials as implemented in GROMACS.

i	$k_i^{(\varphi)}$	$\delta_i^{(\varphi)}$	$k_i^{(\psi)}$	$\delta_i^{(\psi)}$
	$[\mathrm{kJ \cdot mol^{-1}}]$	[rad]	$[\mathrm{kJ \cdot mol^{-1}}]$	[rad]
1	13.97	2.021	-2.812	1.168
2	7.865	-0.978	7.288	1.852
3	-6.49	3.141	2.629	0.091
4	1.229	1.455	0.446	-0.416
5	0.366	-1.894	0.410	1.134
6	0.189	2.629	-0.188	2.96

In order to remove the barriers to backbone torsion and thus to enhance the conformational sampling during simulations, the biasing-potential needs to be subtracted from the standard FF backbone dihedral potential:

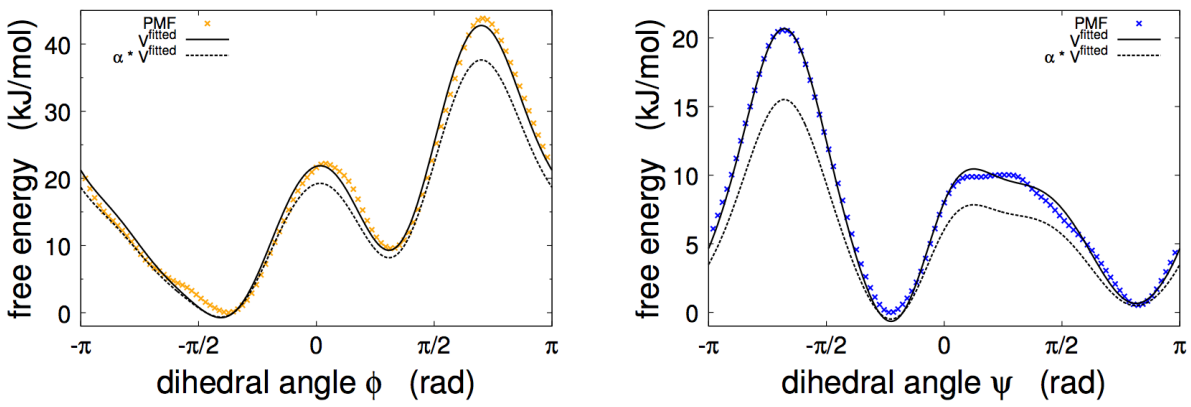

FIG. 1: Comparison of the fitted potentials (V^{fitted}) to the PMFs of the peptide backbone dihedral angles φ and ψ.

$$V^{\text{modified}}(\varphi) \;=\; V^{\text{FF}}(\varphi) - V^{\text{bias}}(\varphi) \tag{2}$$

However, in order to ensure that the low energy regions of the PMF remain favourable during conformational sampling, the fitted functions are not used directly as biasing potentials. Instead, they are rescaled by a factor smaller than one [14]:

$$V^{\text{bias}}(\varphi) \;=\; \alpha \cdot V^{\text{fitted}}(\varphi) \tag{3}$$

The fitted potentials of both dihedral angles φ and ψ are rescaled according to equation 3, with a factor α of 0.88 for φ and 0.75 for ψ in the most strongly biased system in the replica-exchange simulation runs. The rescaled fitted functions $\alpha \cdot V^{\text{fitted}}(\varphi)$ are shown in figure 1. This rescaling ensures that the highest energy barrier of the resulting dihedral potential $V^{\text{modified}}(\varphi)$ (eq. 2) is reduced to approximately 5 kJ/mol. The remaining maximum barrier height of 5 kJ/mol corresponds to nearly 2 $k_{\text{B}}T$, which is a barrier that can be easily overcome during simulations at a temperature of 300 K.

Figure 2 shows the standard dihedral potentials V^{FF} of the GROMOS FF together with the modified potentials V^{modified} that were obtained *via* equation 2.

The BP-REMD simulations were executed employing the H-REMD utilities implemented in GROMACS 4.6. In GROMACS, the potentials of bonded interactions can be interpolated smoothly from state A ($\lambda = 0$) to state B ($\lambda = 1$) using the λ-dependence of potentials. The resulting potential of dihedral angle φ of any interpolated state reads [2, 3]:

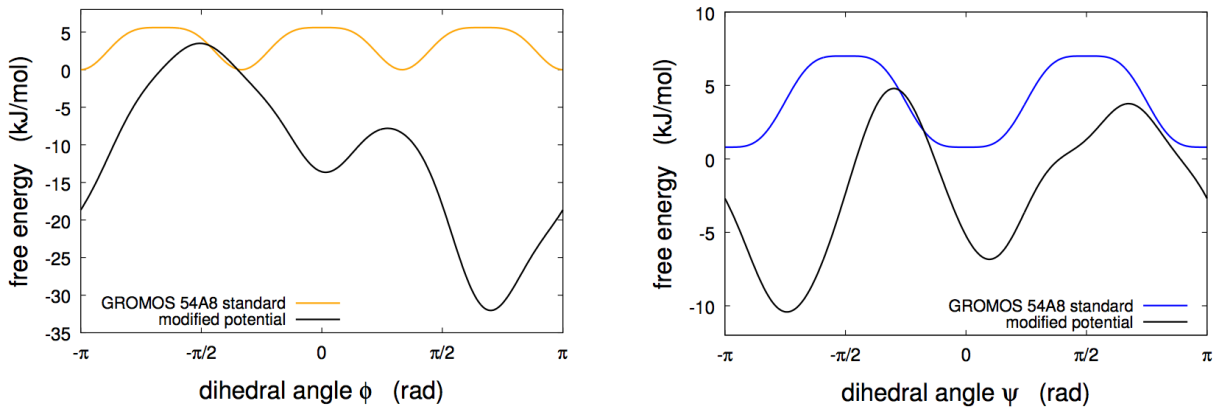

FIG. 2: Comparison of the standard dihedral potentials V^{FF} of the GROMOS 54A8 FF and the modified potentials V^{modified} of the two peptide backbone dihedral angles φ and ψ.

$$V(\varphi) \;=\; [(1-\lambda)k^A + \lambda k^B] \cdot (1 + \cos[n\varphi - (1-\lambda)\delta^A - \lambda\delta^B]) \tag{4}$$

The parameter λ can be assigned to any value between 0 and 1, while k^A, k^B, δ^A and δ^B represent the respective constants of the dihedral potentials of state A and state B. The same function holds for the dihedral angle ψ.

C. Convergence of BP-REMD simulations

The exchanges between replicas and the sampling of replica space was evaluated for all four BP-REMD simulations. It was found to be good in the simulation of 5Glu without calcium ions and reasonable for the simulations of 5Glu with ions and 10Glu without ions, even though further sampling in the latter two simulations would have been advantageous. In the case of 10Glu with calcium ions, a complete random walk in replica space had not yet been achieved. Nevertheless all BP-REMD simulations can be viewed as a success, substantially more efficient sampling of conformational phase space had been achieved than in brute force MD simulations of the systems. The evolution of the histogram of the radius of gyration of 5Glu without ions (figure 3, left panel) shows that the sampling is close to convergence after 400 ns of BP-REMD. These findings can be compared directly with the results of the respective brute-force simulation. After 1000 ns of brute-force simulation, the histogram of the radius of gyration shows no significant changes any more (fig. 3, middle panel). However, this does not mean that this simulations has converged in the sense that the phase space has been sampled comprehensively. A comparison to the results of the BP-REMD simulation (fig. 3, right panel) reveals that there is a region in conformational phase space, with a radius of gyration of pentaGLU of approximately 0.5 nm, which has not yet been sampled during the 2000 ns brute-force simulation. This is remarkable as out of the four systems considered here, this is the system in which comprehensive conformational sampling is most easily achieved. In the presence of ions, the differences between brute-force and BP-REMD simulations in the efficiency of phase space sampling are even more striking.

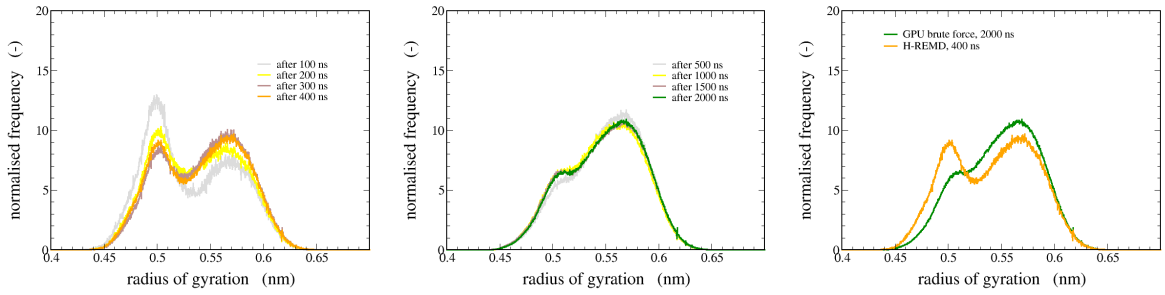

FIG. 3: Evolution of the radius of gyration of 5Glu without calcium ions. Left panel: BP-REMD; Middle panel: Brute force MD; Right panel: comparison of BP-REMD and brute force MD.

D. Sketch-map settings

40,000 data points from each of the four simulations (5Glu and 10 Glu with and without Ca^{2+} ions) were selected randomly. These data points are snapshots from the trajectories and thus contain one configuration of the peptide. As collective variables (CV) which characterise the high-dimensional data points the peptide backbone dihedral angles were used (8 dihedral angles in the case of the 5Glu and 18 dihedral angles in the case of the 10Glu. These span the high-dimensional phase space which is to be projected on two dimensions with the help of sketch-map. The sketch-map analyses were performed with the code provided on the sketch-map webpage according to the detailed instructions of the tutorial on the webpage ([15]). The first step of the sketch-map procedure is the *analysis of pairwise distances* R_{ij} of the high-dimensional data to determine the typical distances between nearby stable conformations ($R_{ij} \approx 3.5$) to be used as key parameter in the adjustment of the sigmoid functions [16]. A subset of 1,000 landmark points was used. As we are interested in a comparison of the conformations of the peptides with and without counter-ions, we used the same projection from high- into low-dimensional space for both data sets. Therefore, 500 landmark-points were selected from each of the two simulations by means of the "farthest point strategy", which guarantees that all the sampled regions of phase space are represented in this selection [15]. The weights of the landmarks were assigned by computing the number of trajectory frames within each landmark's Voronoi polyhedron in the high-dimensional space [16].

E. Sampling of conformation space: 10Glu with and without ions

Also for 10Glu, the presence of calcium ions severely affects the extent of the sampled conformation space:

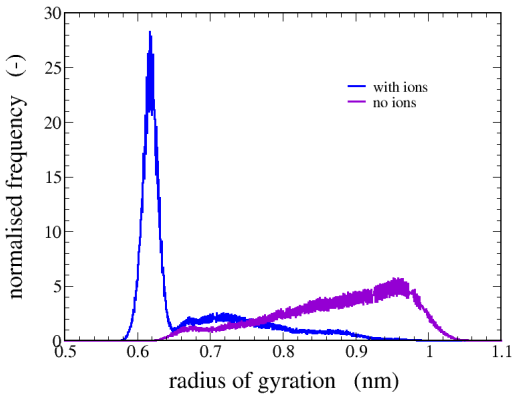

FIG. 4: Distribution of the radius of gyration of 10Glu during the BP-REMD simulations with and without counter-ions.

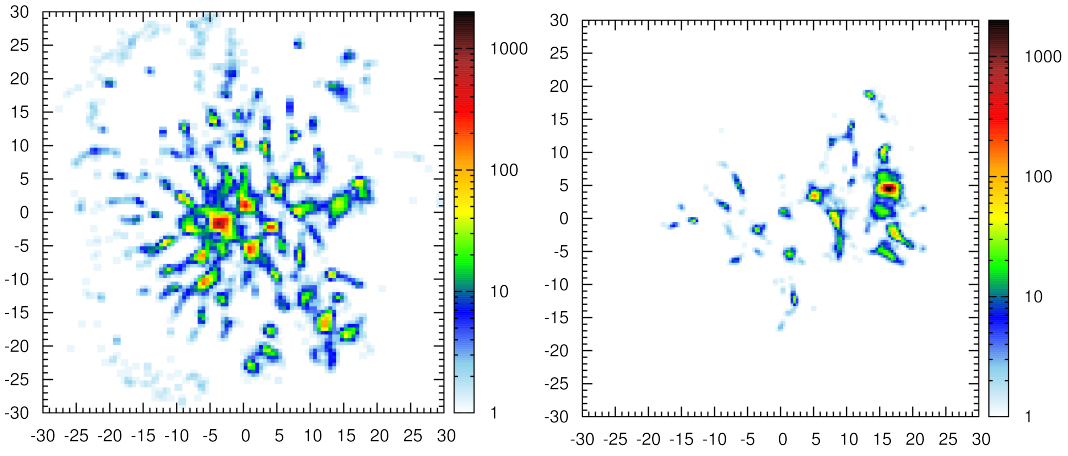

FIG. 5: Two-dimensional sketch-map projections of 40,000 configurations from each the BP-REMD simulations of 10Glu without (left panel) and with calcium ions (right panel). The clustering is based on the similarity of the peptide's backbone conformation and the various regions of the projection represent different types of structures of the ion–peptide complexes. The colour scheme reflects the absolute probability of data points per area.

[1] V. Fischer, K. Landfester and R. Muñoz Espí, *Crystal Growth & Design*, 2011, **11**, 1880–1890.

[2] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *Journal of computational chemistry*, 2005, **26**, 1701–18.

[3] B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *Journal of Chemical Theory and Computation*, 2008, **4**, 435–447.

[4] M. M. Reif, P. H. Hünenberger and C. Oostenbrink, *Journal of chemical theory and computation*, 2012, **8**, 3705–3723.

[5] J. Kahlen, L. Salimi, M. Sulpizi, C. Peter and D. Donadio, *The Journal of Physical Chemistry B*, 2014, **118**, 3960–3972.

[6] W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, 1987.

[7] H. Berendsen, J. Grigera and T. Straatsma, *Journal of Physical Chemistry*, 1987, **91**, 6269–6271.

[8] G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, 014101.

[9] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577.

[10] S. Kannan and M. Zacharias, *Proteins: Structure, Function, and Bioinformatics*, 2007, **66**, 697–706.

[11] K. Ostermeir and M. Zacharias, *Biochimica et Biophysica Acta*, 2013, **1834**, 847–53.

[12] A. Barducci, G. Bussi and M. Parrinello, *Physical Review Letters*, 2008, **100**, 020603.

[13] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Computer Physics Communications*, 2014, **185**, 604–613.

[14] T. P. Straatsma and J. A. McCammon, *The Journal of Chemical Physics*, 1994, **101**, 5032.

[15] `http://sketchmap.berlios.de/compose.php?page=tuts`.

[16] M. Ceriotti, G. A. Tribello and M. Parrinello, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 13023–8.